ABSTRACT

An apparatus and method are provided to recover the high level structure of a program, such as a television or video program using an unsupervised clustering algorithm in concert with a human analyst. The method is comprised of three phases, a first phase, referred to herein as a text type clustering phase, a second phase of genre/sub-genre identification phase in which the genre/sub-genre type of a target program is detected and a third and final phase, referred to herein as a structure recovery phase. The structure recovery phase relies on graphical models to represent program structure. The high level structure of a program, once recovered, may be advantageously used in a recover further information including, but not limited to, temporal events, text events, program events and the like.